LA-UR-80-2485

MASTER

TITLE: Statistical Relationship Between Median Visibility and Conditions of Worst Case Man-Made Impact on Visibility

AUTHOR(S): John D. Gins, Technology Service Corporation
David Nochumson, S-3
John Trijonis, Santa Fe Research Corporation

SUBMITTED TO: Symposium on Plumes and Visibility
held at Grand Canyon November 10-14, 1980.

DISCLAIMER

University of California

LOS ALAMOS SCIENTIFIC LABORATORY
Post Office Box 1663 Los Alamos, New Mexico 87545
An Affirmative Action/Equal Opportunity Employer

Form No. 836 R3
St. No. 2629
12/78

UNITED STATES
DEPARTMENT OF ENERGY
CONTRACT W-7405-ENG. 36

STATISTICAL RELATIONSHIP BETWEEN MEDIAN VISIBILITY AND CONDITIONS OF WORST-CASE MANMADE IMPACT ON VISIBILITY

John D. Gins
Technology Service Corporation, Santa Monica, California

David H. Nochumson
Los Alamos Scientific Laboratory, Los Alamos, New Mexico

John Trijonis
Santa Fe Research Corporation, Santa Fe, New Mexico

Abstract - A study was conducted of the statistical relationships between median visibility and the levels of visibility associated with worst-case manmade impacts. The data base for the study consisted of midday visibility recordings for the years 1974-1976 at 28 suburban/nonurban airports throughout the United States. The visibility recordings were converted to estimates of extinction coefficients with the use of the Koschmeider formula. The data were sorted according to meteorology in order to eliminate days that were obviously dominated by natural causes of poor visibility. Three approaches were used for relating worst-case (90th through 99th percentile) extinction to median extinction. The first approach was based upon frequency distribution functions. The second used observed ratios of upper percentile to median extinction. The third employed regression techniques. All of the relationships were formulated and evaluated with the 1974-1976 data on a national/annual basis as well as regional/quarterly basis. Performance tests were also run against 1954-1956 data at 11 of the 28 sites. Simple ratio relationships are recommended for use in translating median visibility impacts into worst-case impacts. The errors associated with these ratio models are approximately 30%, which is less than the error typically associated with mathematical dispersion models.

1. INTRODUCTION

Los Alamos Scientific Laboratory (LASL) has developed and tested a methodology for estimating regional-scale visibility (light extinction and visual range) from regional-scale aerosol concentrations, relative humidities, and air pollutant emissions. (Nochumson, Wecksung, and Gurule, 1979a and 1979b). This methodology has been used for evaluating the visibility impacts of future energy scenarios; it yields predictions for indices of median (50th percentile) visibility. There is a need to develop relationships for translating these median estimates into short-term predictions for worstcase days (e.g., 90th or 99th percentiles of manmade impacts). This paper presents relationships which were developed through the analysis of statistical distributions of airport visibility data.

In the development of the relationships between median and worst-case (upper percentile visibility), the following four steps of analysis were followed.

(1) Selection of visibility observation sites, years of observations, and time of day of observations.

(2) Screening of data according to meteorology, attempting to remove days heavily influenced by natural phenomena (e.g., fog, precipitation, blowing dust, or low clouds).

(3) Using 1974-1976 data, development of relationships for translating worst-case (90th or 99th percentile) extinction to median extinction. Three approaches were followed to develop the relationships. One was based upon frequency distribution functions. A second was based upon observed ratios of upper percentile to median extinction. A third employed regression techniques.

(4) Evaluation of the performance of the relationships regionally as well as nationally, and quarterly as well as annually. Testing of models for some sites with the use of 1954-1956 data.

Steps (1) and (2) are discussed in Section 3.0 of this paper. Steps (3) and (4) are discussed in Section 4.0 of this paper. Basic concepts and definitions fundamental to this analysis are discussed in Section 2.

2. BASIC CONCEPTS AND DEFINITIONS

Visibility refers to the clarity of the atmosphere and can be defined quantitatively in terms of discoloration (wavelength shifts in light produced by the atmosphere), contrast (the relative brightness of visible objects), and/or visual range (the farthest distance that one would be able to distinguish a large black object against the horizon sky). Because this study is based on airport weather station measurements of visual range, we define visibility as visual range and use the two terms interchangeably. It should be noted that the concept of visual range makes the most sense in situations of large-scale homogenous haze, which is the type of visibility phenomenon addressed in this paper.

Visibility through the atmosphere is restricted by the absorption and scattering of light by both gases and particles. The sum of absorption and scattering is called total extinction which is measured by the extinction coefficient "B". The extinction coefficient represents the fraction of light that is attenuated per unit distance as a light beam traverses the atmosphere. In a homogeneous atmosphere, visibility is inversely proportional to extinction. For a standard observer (one able to perceive a 2% contrast), the Koschmeider formula expressing this relationship is:

$$B = 24.3/V \qquad (1)$$

where the units of visibility or visual range, V are miles, and the units of B are 10^{-4} meters^{-1}, the standard units for extinction.

The analyses in this study were actually conducted with extinction data derived from the visibility data according to Equation (1). When examining worst-case impacts, there is an advantage in using extinction rather than visual range. It is known that, with other factors held constant, each component of extinction (scattering by gases, absorption by gases, scattering by particles, and absorption by particles) is directly proportional to the concentration of gases or particles. Thus, we would expect the statistical distribution of extinction to resemble the well-studied statistical distribution of pollutant concentrations.

As remarked earlier, the data base for this study consisted of airport weather station measurements of visual range. An important aspect of airport visibility measurements is their discrete nature; visual range is generally reported with respect to a discrete set of visibility markers (e.g., mountains, buildings, or towers). Trijonis and Yuan (1978) and Trijonis (1979) have emphasized that, because of the nature of reporting practices, airport data should only by plotted as cumulative frequency in a certain order (starting with the farthest visibility marker). This and other problems introduced by the discrete nature of the data are discussed further in this paper.

3. DESCRIPTION OF THE DATA BASE

This section discusses the selection of airport sites, observation times, and meteorological screening procedures. A brief description of the spatial and temporal patterns in median visibility is also presented.

3.1 Selection of Sites and Observation Times

Visibility observations for the years 1974-1976 at 28 airport weather stations served as the data base for the study. The criteria for selection of the 28 stations were: (1) that the sites were suburban/nonurban rather than urban, (2) that the median visual range was less than, or about the same as, the distance to the farthest visibility marker, (3) that no major site relocations or changes in reporting practices occurred in 1974-1976, and (4) that the stations were well spread throughout the continental United States. The 28 study locations are displayed in Figure 1 and listed in Table 1. As shown in Figure 1, the sites were grouped according to four major regions: Pacific, Mountain, Central, and Eastern.

Visual range was measured in miles with respect to a system of fixed markers. As an example, for a station with farthest markers at 50 and 75 miles, a recording of 50 miles implies only that visual range is somewhere between 50 and 75 miles. Because of the nature of visibility reporting practices at airports, it is important that the visibility frequency distribution be plotted in the proper way with respect to the marker system (Trijonis and Shapland 1979; Trijonis 1979; Trijonis and Yuan 1978). In this study, the marker system listed in Table 1 was used for all visibilities above 3 miles. The few nonstandard readings that were sometimes recorded at the airports were reassigned according to this marker system (e.g., referring to the previous example, an occasional recording of 55 miles was reassigned to the 50-mile marker).

The visibility observations and other weather parameters were extracted from National Climatic Center TDF-14 data tapes. These tapes contain recordings for every third hour, Greenwich Mean Time. Only the single nearest-noon observation (13:00 Eastern and Pacific Time, noon Central Time, and 11:00 Mountain Time) was used in the study. We did not use all four daylight hours separately because diurnal visibility fluctuations would then be mixed with day-to-day visibility fluctuations (our real interest). Also, because of the need to summarize the visibility frequency distributions according to a discrete set of markers, it was not appropriate to average the four daylight observations.

Data for 1954-1956 at 11 of the sites were used to check the relationships formulated with the 1974-1976 data. These 11 sites were chosen

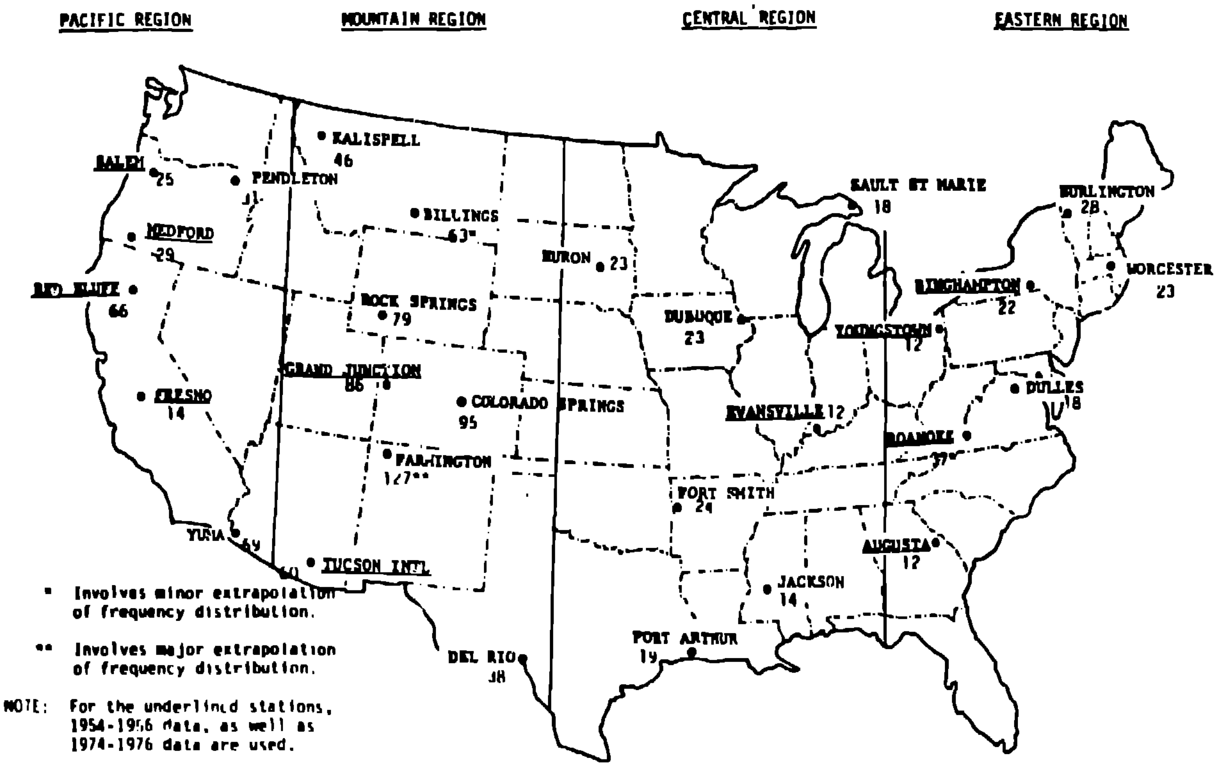

Fig. 1. Map of the 28 study locations including the median (weather-sorted) visibilities for 1974-1976.

TABLE 1. LOCATIONS, TIME PERIODS, AND VISIBILITY MARKERS FOR THE STUDY SITES.

Location	Time Period	Markers (Miles)
Pacific		
Fresno, Calif.	54-56 & 74-76	30 20 15 10 7 5 4 3
Medford, Ore.	54-56 & 74-76	30 25 20 15 10 7 5 4 3
Pendleton, Ore.	74-76	40 30 25 20 15 10 7 5 4 3
Red Bluff, Calif.	54-56 & 74-76	85 50 30 15 10 7 5 3
Salem, Ore.	54-56 & 74-76	75 50 40 30 20 15 10 7 5 4 3
Yuma, Ariz.	74-76	85 50 30 20 15 10 5 3
Rocky Mountain		
Billings, Mont.	74-76	60 50 40 30 20 10 7 5 4 3
Colorado Springs, Col.	74-76	100 65 45 25 10 5 3
Del Rio, Texas	73-75	75 50 30 15 10 7 5 4 3
Farmington, N.M.	74-76	75 50 40 30 20 10 7 6 5 4 3
Grand Junction, Col.	54-56 & 74-76	90 80 70 60 50 40 30 20 10 5 3
Kalispell, Mont.	74-76	50 40 30 20 15 10 7 6 5 4 3
Rock Springs, Wyo.	74-76	90 60 30 15 10 7 5 3
Tucson, Ariz.	54-56	70 60 50 40 30 20 15 10 7 5 3
Tucson, Ariz.	74-76	60 50 40 30 20 15 10 7 5 3
Central		
Dubuque, Iowa	74-76	30 20 10 7 6 5 4 3
Evansville, Ind.	54-56 & 74-76	15 10 7 6 5 4 3
Fort Smith, Ark.	74-76	25 20 15 10 7 5 4 3
Huron S.D.	74-76	30 15 10 7 5 4 3
Jackson, Miss.	74-76	20 15 12 10 7 5 4 3
Port Arthur, Texas	74-76	20 15 10 7 5 4 3
Sault St. Marie, Mich.	74-76	20 15 10 7 5 4 3
Eastern		
Augusta, Ga.	54-56 & 74-76	15 10 7 5 4 3
Binghamton, Mass.	54-56 & 74-76	45 20 15 10 7 5 4 3
Burlington, Ver.	74-76	45 30 25 20 15 10 7 5 4 3
Dulles, Vir.	74-76	25 20 15 12 10 7 5 4 3
Roanoke, Vir.	54-56 & 74-76	35 25 15 10 7 5 4 3
Worcester, Mass.	74-76	40 30 25 20 15 10 7 6 5 4 3
Youngstown, Ohio	54-56 & 74-76	25 15 10 7 5 3

according to the criterion that there were no major relocations or reporting changes from 1954 to 1976.

3.2 Meteorological Screening

Visibility reduction involves both natural contributions (e.g. Rayleigh scatter by air molecules, or natural aerosols such as water droplets or dust) and anthropogenic contributions (e.g. manmade aerosols). It is the purpose of this report to consider worst-case manmade visibility impacts. Although it is impossible to entirely segregate manmade impacts from natural impacts, the data were sorted by meteorological parameters in an attempt to eliminate days when natural factors dominated.

Any day with a midday recording of fog, precipitation, blowing dust, or blowing snow was eliminated from the data base. Also, a study was conducted of the relationship between visibility and various weather parameters in order to identify other situations that might be dominated by natural processes. Based on this study, we decided to also eliminate days when wind speed exceeded 25 knots (worst-case visibility on such days would more likely be due to dust than to manmade sources), or when ceiling height was less than 1500 feet (clouds might be obscuring visibility markers on such days).

3.3 Spatial and Temporal Patterns

Before proceeding with the statistical analysis effort, it is useful to give the reader a feel for the data by examining the spatial and temporal patterns of median visibility. Median visibilities for the weather-sorted data are computed using interpolation (or extrapolation) procedures based on the exponential distribution.

Figure 1 illustrates the geographical pattern of median visibilities for the weather-sorted data at the 28 study locations. The spatial patterns are similar to those found by Trijonis and Shapland (1979) using data not sorted for meteorology from 101 locations. The major similarities are that visibility is best in the Rocky Mountain Southwest and worst east of the Mississippi River.

Table 2 compares median visibilities for the weather-sorted data to median visibilities for all data. As expected, the weather-sorted medians are higher than the medians for all the data, generally about 5 to 15% higher in the Pacific Region, 5 to 10% higher in the Mountain Region (the only major exception being in Farmington, NM, where the weather-sorted median may be artificially high due to the extrapolation procedure), 5 to 20% higher in the Central Region, and 10 to 25% higher in the Eastern Region. These figures suggest that weather-related reductions in visibility tend to be least in the Mountain Region; this is not unexpected, considering the arid climate of that region.

The seasonal patterns in median weather-sorted visibilities were also investigated. In agreement with previous findings by Trijonis and Yuan (1978) and Husar et al. (1979), the seasonal patterns in the Central and Eastern Regions displayed a summer (third quarter) minimum in visibility. In agreement with the findings of Roberts et al. (1975), most of the Mountain Region sites exhibited minimal visibilities during the spring and summer. Most of the Pacific Region sites showed a springtime maximum in visibility with lowest visibilities in fall and winter.

For the 11 sites where 1954-1956 data as well as 1974-1976 data were studied, the visibility changes from 1954-1956 to 1974-1976 were investigated. The sites in the Pacific Region showed mixed trends from 1954-1956 to 1974-1976, although substantial decreases in reported visibilities occurred at Fresno and Salem. The two Mountain Region sites showed slight decreases from 1954-1956 to 1974-1976; this is consistent

TABLE 2. COMPARISON OF MEDIAN VISIBILITIES FOR WEATHER-SORTED DATA AND ALL DATA.

Location	Median Mid-day Visibility for all Data in 1974-1976 (Trijonis and Shapland, 1979)	Median Mid-day Visibility for Weather-Sorted Data in 1974-1976	Percent Change (Weather Adjusted Data Minus All Data)
Pacific			
Fresno	13 miles	14 miles	+8%
Medford	25	29	+16%
Pendleton	29	31	+7%
Red Bluff	61	6[illegible]	+8%
Salem	23	25	+9%
Yuma	59	59	0%
Rocky Mountain			
Billings	60	63	+5%
Colorado Springs	90	95	+5%
Del Rio	34	38	+12%
Farmington	80	127	+59%
Grand Junction	84	86	+2%
Kalispell	41	46	+12%
Rock Springs	76	79	+4%
Tucson Int'l	60	60	0%
Central			
Dubuque	19	23	+21%
Evansville	10	12	+20%
Fort Smith	22	24	+9%
Huron	20	23	+15%
Jackson	13	14	+8%
Port Arthur	18	19	+6%
Sault St Marie	17	18	+6%
Eastern			
Augusta	11	12	+9%
Binghampton	17	22	+29%
Burlington	23	28	+22%
Dulles	17	18	+6%
Roanoke	31	37	+19%
Worchester	18	23	+28%
Youngstown	10	12	+20%

with earlier findings (Trijonis 1979; Marians and Trijonis 1979; Latimer et al. 1978) that visibility decreased in the Rocky Mountain West from the mid-1950s to the early 1970s and then improved somewhat from the early 1970s to the middle 1970s. Four of the five sites in the Central and Eastern Regions exhibited decreasing trends in agreement with the increasing trend in haziness reported by Trijonis and Yuan (1978), and Husar et al. (1979).

4. STATISTICAL METHODS

Three approaches were followed to develop relationships between median and upper (90th to 99th) percentile visibility. The first approach was to fit the data to several candidate frequency distribution functions. The second approach, the simplest one, was based on actual ratios between upper percentile and median extinction. The third approach employed regression equations to relate upper percentile to median visibility. For each approach, relationships were developed on an annual as well as quarterly basis. For the latter two approaches, relationships were developed on a national as well as a regional basis. The development of the three approaches and an assessment of their performance is discussed in the following subsections.

4.1 Frequency Distribution Functions

Knowledge of the underlying mathematical distributions that approximate extinction data is very useful in formulating statistical models. The purpose of this section is to briefly explore if there is a distribution function that characterizes extinction data. There are several characteristics of the extinction data that must be taken into account. The data are derived from airport visibility data which are given for distinct markers rather than for continuous readings. Thus, one extinction value really represents an interval of extinction values. This property is called grouping. We use the Chi-square goodness-of-fit test to take advantage of the grouping of the data because it requires the data first be broken into discrete cells (Breiman, 1973). The second characteristic of extinction data is that an extinction coefficient of zero, which corresponds to infinite visibility, does not occur because of the lower limit on extinction imposed by Rayleigh scattering; and possibly because of the contribution by background aerosols. This indicates that the possible distributions must either take into account the shift away from zero, or the extinction data must be shifted back to zero by some appropriate transformation.

The cumulative frequency distribution of the extinction data at each site was plotted on exponential distribution paper on an annual as well as a quarterly basis. Figure 2 is a plot of the annual frequency distribution for Dubuque. The data in most of these plots tend to follow a straight line over the area of interest (from the 50th to the 99th percentiles) with substantial bends in some tails beyond the 99th percentile. Straight lines on exponential paper indicate that the data might fit an exponential distribution but it is wise to perform goodness-of-fit tests to verify the graphical evidence.

The Weibull, lognormal, gamma and exponential distributions were fitted to the extinction data because they are commonly used "single tailed" distributions. The density functions of the various distributions and the estimators of their parameters are well known and are available from a number of standard statistical texts. The two-parameter shifted exponential distribution was used. The second parameter is a location parameter that accounts for the shift of the extinction data away from zero. Because the extinction data are already broken into distinct groups and because the Chi-square test requires that the data be broken into distinct cells, we used the Chi-square goodness-of-fit test to evaluate the fit of various distributions.

The Chi-square goodness-of-fit statistic is the sum over all cells of

$$\frac{(O(i)-E(i))^2}{E(i)} \tag{2}$$

where O(i) is the observed number in cell i and E(i) is the expected number for cell i. In order to obtain a good Chi-square approximation E(i) must be greater than or equal to 5. If this does not occur, then cells will have to be combined (Breiman, 1973). The degrees of freedom associated with the Chi-square statistic is the number of cells minus one minus the number of parameters estimated for the distribution being tested.

The goodness-of-fit for the four two-parameter distributions was first tested on the annual data using a 95% significance level as the rejection criterion. The tests were conducted over the subset of each data set that included the 50th through 99th percentiles. The gamma, lognormal, and the Weibull were rejected for all 39 site-years and the exponential distribution was rejected for all but 5 of the data sets (see Table 3). Since the exponential distribution did

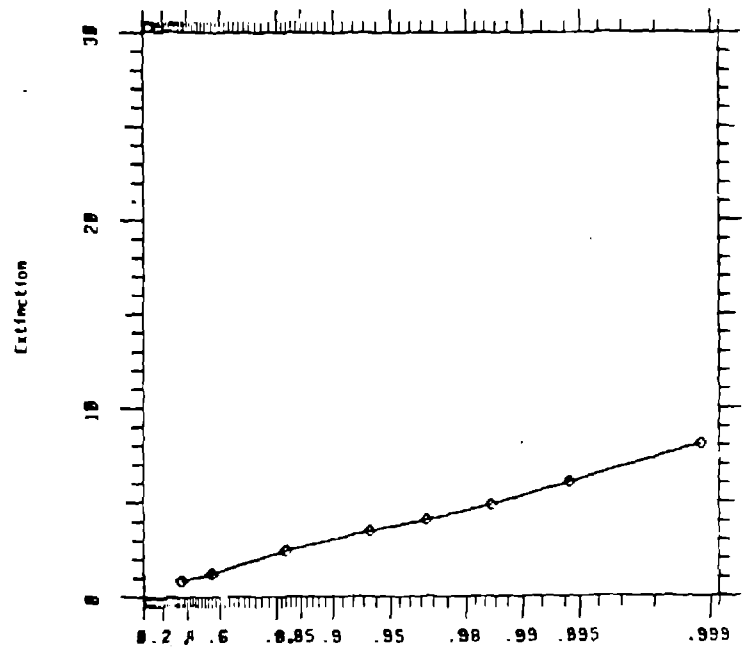

Fig. 2. Exponential plot of extinction at Dubuque for 1974-1976.

TABLE 3. CHI-SQUARE GOODNESS-OF-FIT TEST RESULTS FOR ANNUAL DATA.

Location	Years	Exponential Rejected?	Weibull Rejected?	Log-normal Rejected?	Gamma Rejected?
Pacific					
Fresno	54-56	Yes	Yes	Yes	Yes
Fresno	74-76	Yes	Yes	Yes	Yes
Medford	54-56	Yes	Yes	Yes	Yes
Medford	74-76	Yes	Yes	Yes	Yes
Pendleton	74-76	Yes	Yes	Yes	Yes
Red Bluff	54-56	Yes	Yes	Yes	Yes
Red Bluff	74-76	Yes	Yes	Yes	Yes
Salem	54-56	Yes	Yes	Yes	Yes
Salem	74-76	Yes	Yes	Yes	Yes
Yuma	74-76	Yes	Yes	Yes	Yes
Rocky Mountain					
Billings	74-76	Yes	Yes	Yes	Yes
Colorado Springs	74-76	Yes	Yes	Yes	Yes
Del Rio	73-75	No	Yes	Yes	Yes
Farmington	74-76	Yes	Yes	Yes	Yes
Grand Junction	54-56	Yes	Yes	Yes	Yes
Grand Junction	74-76	Yes	Yes	Yes	Yes
Kalispell	74-76	Yes	Yes	Yes	Yes
Rock Springs	74-76	Yes	Yes	Yes	Yes
Tucson	54-56	Yes	Yes	Yes	Yes
Tucson Intl	74-76	Yes	Yes	Yes	Yes
Central					
Dubuque	74-76	Yes	Yes	Yes	Yes
Evansville	54-56	No	Yes	Yes	Yes
Evansville	74-76	Yes	Yes	Yes	Yes
Fort Smith	74-76	Yes	Yes	Yes	Yes
Huron	74-76	Yes	Yes	Yes	Yes
Jackson	74-76	Yes	Yes	Yes	Yes
Port Arthur	74-76	No	Yes	Yes	Yes
Sault St Marie	74-76	Yes	Yes	Yes	Yes
Eastern					
Augusta	54-56	Yes	Yes	Yes	Yes
Augusta	74-76	Yes	Yes	Yes	Yes
Binghampton	54-56	No	Yes	Yes	Yes
Binghampton	74-76	Yes	Yes	Yes	Yes
Burlington	74-76	Yes	Yes	Yes	Yes
Dulles	74-76	Yes	Yes	Yes	Yes
Roanoke	54-56	Yes	Yes	Yes	Yes
Roanoke	74-76	Yes	Yes	Yes	Yes
Worcester	74-76	Yes	Yes	Yes	Yes
Youngstown	54-56	Yes	Yes	Yes	Yes
Youngstown	74-76	No	Yes	Yes	Yes

better than the other distributions, the exponential distribution was used to fit quarterly as well as annual data sets. The exponential distribution fit better on a quarterly basis than on the annual basis. Over the four quarters, 11 were not rejected for the first quarter, 14 were not rejected for the second, 16 were not rejected for the third, and 8 were not rejected for the fourth. The Rocky Mountain Region did the worst in the first, second, and third quarters. The Central Region did the worst in the fourth quarter. The 1974-76 period had six fewer rejections than the 1954-56 period for the comparable sites.

Because the mathematical distribution functions did not fare well in the goodness-of-fit tests, we decided not to propose statistical relationships based solely on a distribution function. We did do some testing of such relationships and found that they did not perform as well as the other two approaches. The exponential distribution did, however, prove useful in one respect: we used it to interpolate between the discrete data points so as to form continuous cumulative frequency distributions.

4.2 Estimating the Median and Other Percentiles

The median and specified upper percentile extinction values were used in the estimation and assessment of the ratio and regression relationships. They were estimated primarily by interpolation from the extinction data which were continuous random variables recorded at discrete points. The classical statistical method of computing the median is to rank the data and choose the middle value as the median. This is not a reasonable way of estimating the median extinction coefficient because of the discrete nature in which airport visibility data are reported. We have chosen to estimate the 50th percentile by interpolation over the interval into which the median would fall. We call this estimate the median. The other percentiles also were obtained by means of interpolation.

The interpolation procedure chosen was a logarithmic interpolation procedure which can be derived from the exponential distribution. For the desired value, X, associated with the percentile 100P(X), the formula is

$$X=(C-B)(\ln(1-P(X))-\ln(1-P(B)))/(\ln(1-P(C))-\ln(1-P(B)))+B \qquad (3)$$

where $0<P(B)\leq P(X)\leq P(C)<1$ and B and C are endpoints of the interval containing X.

For a few sites, the 50th percentile had to be extrapolated from the data because the lowest extinction coefficient accounted for more than 50% of the data. The median X(.5) was derived for such cases by the following equation:

$$X(.5)=(C-\alpha)(\ln(.5))/(\ln(1-P(C)))+\alpha \qquad (4)$$

where α is the location parameter estimated for the two-parameter shifted exponential distribution function.

4.3 The Ratio Method

Average ratios of the upper percentiles to the median are estimated for each region and the nation (see Table 4). Extinction data from the 1974-76 time period were used to calculate the averages, and the data from the 1954-56 time period were used to assess the performance of the ratio method (see Subsection 4.5). The ratios measure, in some sense, the "episodicity" of extinction (i.e., the magnitude of worst-case

TABLE 4. AVERAGE RATIOS OF THE MEDIAN TO SELECTED PERCENTILES OF THE EXTINCTION COEFFICIENT.

Period	Percentile	National	Pacific	Rocky Mountain	Central	Eastern
Annual	90	2.40	2.50	2.25	2.04	2.95
	91	2.51	2.64	2.37	2.11	2.95
	92	2.63	2.80	2.50	2.18	3.06
	93	2.77	3.01	2.65	2.27	3.22
	94	2.95	3.26	2.82	2.36	3.40
	95	3.15	3.56	3.03	2.48	3.60
	96	3.41	3.94	3.33	2.62	3.84
	97	3.79	4.49	3.79	2.6[illegible]	4.16
	98	4.39	5.25	4.49	3.08	4.85
	99	5.63	6.50	5.88	3.60	5.83
First Quarter	90	2.51	2.70	3.01	1.99	2.32
	91	2.62	2.80	3.15	2.06	2.41
	92	2.73	2.92	3.30	2.13	2.50
	93	2.86	3.05	3.50	2.21	2.60
	94	3.02	3.25	3.74	2.30	2.71
	95	3.22	3.50	4.03	2.40	2.88
	96	3.49	3.82	4.44	2.51	3.10
	97	3.91	4.38	5.12	2.66	3.37
	98	4.51	5.07	6.17	2.87	3.80
	99	5.57	6.34	7.74	3.57	4.45
Second Quarter	90	2.27	2.02	2.15	2.17	2.74
	91	2.36	2.07	2.25	2.24	2.85
	92	2.45	2.12	2.36	2.33	2.95
	93	2.56	2.19	2.50	2.42	3.09
	94	2.69	2.27	2.65	2.54	3.25
	95	2.85	2.36	2.84	2.67	3.44
	96	3.04	2.48	3.09	2.84	3.65
	97	3.28	2.62	3.45	3.00	3.94
	98	3.67	2.81	3.97	3.27	4.47
	99	4.42	3.14	5.24	3.80	5.21
Third Quarter	90	2.18	2.05	1.81	2.08	2.81
	91	2.26	2.13	1.89	2.16	2.91
	92	2.35	2.21	1.97	2.24	3.03
	93	2.46	2.31	2.05	2.33	3.17
	94	2.59	2.42	2.16	[illegible].43	3.37
	95	2.75	2.56	2.30	2.[illegible]6	3.61
	96	2.94	2.71	2.46	2.72	3.89
	97	3.16	2.87	2.66	2.90	4.23
	98	3.47	3.08	3.01	3.13	4.67
	99	4.04	[illegible].44	3.67	3.50	5.51
Fourth Quarter	90	2.54	3.17	2.68	1.89	2.47
	91	2.64	3.32	2.81	1.95	2.56
	92	2.76	3.49	2.96	2.00	2.66
	93	2.89	3.68	3.14	2.07	2.76
	94	3.05	3.92	3.35	2.14	2.68
	95	3.23	4.18	3.60	2.22	3.01
	96	3.47	4.51	3.97	2.30	3.18
	97	3.88	4.97	4.77	2.41	3.40
	98	4.45	5.50	5.91	2.56	3.79
	99	5.46	6.45	7.66	2.83	4.49

conditions compared to median conditions). The ratios vary widely from site to site with the highest ratios tending to occur in the Eastern Region and lowest in the Central Region. The Rocky Mountain and Pacific Regions tend to have their highest ratios during the first and fourth quarters, while the Eastern and Central Regions have their highest ratios during the second and third quarters. Of the 11 sites with two time periods of data, all but Roanoke, Salem, Fresno, and Medford show nearly the same annual ratios over both time periods.

4.4 The Regression Approach

Regression analysis was performed to estimate the relationship between the median and the upper percentile extinction. Regression equations were estimated on a national as well as on a regional basis. Data from the 1974-76 time period were used in the estimation of the regression model parameters, and data from the 1954-56 time period were used to assess the performance of the regression model.

The general model to be examined was

$$X = aX(.5)^b(-\ln(1-P))^c \quad (5)$$

where X is the extinction associated with a percentile, P, X(.5) is the median extinction, and a, b, and c are model parameters to be estimated. The percentiles, P, were transformed in order to simulate the increase in the magnitude of the extinction, X, as the associated percentile, P, increases. The particular transformation used is reflective of the exponential distribution. Thus, if the associated percentile is P, then the transformed percentile is $-\ln(1-P)$.

In order to fit this model, natural logarithms of the dependent and independent variables were taken and a least-squares linear regression was performed on the regression parameters in order to get equations in the form of equation 5. The dependent variable, X, was limited to the observed values that lay between the 90th and 99th percentiles or to those observations that were used in the interpolation of the 90th or 99th percentiles. The national/regional, annual/quarterly equations are given in Table 5 along with the percent variation explained (PVE) by the regression and the number of percentiles (cases) used in the estimation of the regression.

For the annual data the fits of the regression equations could be considered good with respect to the PVE for all regions and nationwide. The regression equations fit poorly with respect to PVE for the first quarter data. For the Central Region in the second quarter and the Eastern Region in the fourth quarter, the regression equations fit well with respect to PVE. The regression equation for the third quarter national data also fits well with respect to PVE. The other regression fits did not perform as well.

4.5 Performance Assessment for the Ratio and Regression Relationships

The purpose of this subsection is to assess the performance of the average ratio and regression relationships for predicting extinction values between the 90th and 99th percentiles. The assessment is conducted with the 1974-76 data used to estimate the relationships and the independent 1954-1956 data set used to test them.

Because of the small number of observations that fall in the 90th to 99th percentile interval, interpolated extinction values were derived for each 0.1th percentile increment, starting at the 90th percentile and going to the 99th percentile. This procedure gives 91 points for each site-year to use in testing the average ratio and regression relationships.

Three measures of error are used to evaluate the results. One measure of error is the average absolute percent error (AAPE) which is given by the following formula:

$$AAPE = \left(\sum_{i=1}^{n} |OBS_i - PRED_i| / OBS_i\right) 100/n \qquad (6)$$

where OBS_i is the observed value and $PRED_i$ is the predicted value. A second measure of error is the average error of the estimate (AEE), which is given by the following formula.

$$AEE = \sum_{i=1}^{n} (OBS_i - PRED_i)/n \qquad (7)$$

A third measure of error is the standard estimation error (SEE) which is given by the following formula.

$$SEE = \left(\sum_{i=1}^{n} (OBS_i - PRED_i)^2/n\right)^{.5} \qquad (8)$$

TABLE 5. EQUATIONS BASED ON THE RESULTS OF REGRESSIONS DONE ON 1974-76 EXTINCTION DATA.

Region	PVE	CASES	EQUATION
			ANNUAL
National	82.7	120	$X(P)=0.941X(.5)^{0.875}(-\ln(1-P))^{1.074}$
Pacific	78.4	29	$X(P)=0.986X(.5)^{0.967}(-\ln(1-P))^{1.150}$
Rocky Mountain	74.9	33	$X(P)=0.517X(.5)^{0.603}(-\ln(1-P))^{1.225}$
Central	87.7	28	$X(P)=1.182X(.5)^{0.498}(-\ln(1-P))^{0.822}$
Eastern	90.3	30	$X(P)=1.374X(.5)^{0.304}(-\ln(1-P))^{0.989}$
			First Quarter
National	40.5	106	$X(P)=2.331X(.5)^{0.580}(-\ln(1-P))^{0.245}$
Pacific	47.2	27	$X(P)=2.739X(.5)^{0.811}(-\ln(1-P))^{0.282}$
Rocky Mountain	5.8	29	$X(P)=2.338X(.5)^{0.264}(-\ln(1-P))^{-.198}$
Central	44.2	25	$X(P)=1.695X(.5)^{0.212}(-\ln(1-P))^{0.512}$
Eastern	46.1	25	$X(P)=1.792X(.5)^{0.376}(-\ln(1-P))^{0.510}$
			Second Quarter
National	65.5	105	$X(P)=1.771X(.5)^{0.860}(-\ln(1-P))^{0.458}$
Pacific	53.3	18	$X(P)=1.320X(.5)^{0.541}(-\ln(1-P))^{0.353}$
Rocky Mountain	19.4	32	$X(P)=1.108X(.5)^{0.488}(-\ln(1-P))^{0.443}$
Central	81.0	26	$X(P)=1.336X(.5)^{0.307}(-\ln(1-P))^{0.790}$
Eastern	17.7	24	$X(P)=3.085X(.5)^{0.166}(-\ln(1-P))^{0\text{[illegible]}10}$
			Third Quarter
National	73.4	110	$X(P)=2.097X(.5)^{1.048}(-\ln(1-P))^{0.299}$
Pacific	48.6	20	$X(P)=1.622X(.5)^{0.597}(-\ln(1-P))^{0.321}$
Rocky Mountain	37.2	31	$X(P)=2.399X(.5)^{1.074}(-\ln(1-P))^{0.052}$
Central	40.6	27	$X(P)=2.026X(.5)^{0.636}(-\ln(1-P))^{0.379}$
Eastern	19.5	32	$X(P)=3.712X(.5)^{0.275}(-\ln(1-P))^{0.3\text{[illegible]}}$
			Fourth Quarter
National	58.9	108	$X(P)=1.670X(.5)^{0.690}(-\ln(1-P))^{0.556}$
Pacific	59.0	24	$X(P)=2.899X(.5)^{0.653}(-\ln(1-P))^{0.304}$
Rocky Mountain	27.6	35	$X(P)=1.\text{[illegible]}16X(.5)^{0.609}(-\ln(1-P))^{0.503}$
Central	63.8	23	$X(P)=1.601X(.5)^{0.421}(-\ln(1-P))^{0.469}$
Eastern	83.4	26	$X(P)=1.299X(.5)^{0.402}(-\ln(1-P))^{0.809}$

For the national/annual relationships, the measures of error are summarized in Table 6. In general, the relationships performed well in terms of their estimation and prediction errors. The average absolute percent error predominantly ranged between 20 to 40%, which compares well with air quality models that frequently give factor-of-two agreement with observation. There were differences in performance between the four types of relationships tested. Overall, the regional relationships tended to perform better than did the national relationships. The regional ratio and the national regression relationships tended to perform better than the other two relationships on the test data sets (1954-1956). The regional regression relationship tended to perform best on the fit data sets (1974-1976). The regional regression relationship tended to fit the data best in the Pacific and Central Regions. The national regression relationship tended to fit the data best in the Eastern Region. All relationships except the national

TABLE 6. SUMMARY FOR THE ANNUAL RELATIONSHIPS.

Location	Years	Error Using Ratios			Error Using Equations			90% to 99%	
		AAPE	AEE	SEE	AAPE	AEE	SEE	Mean	SD*
Pacific									
Fresno	54-56	5.46	-0.26	0.34	13.35	-0.70	0.74	5.07	1.14
Fresno	74-76	26.60	-2.21	2.38	34.56	-2.88	3.10	8.10	2.24
Medford	54-56	33.77	-1.54	1.66	35.48	-1.63	1.77	4.43	1.28
Medford	74-76	9.09	-0.24	0.45	9.95	-0.30	0.55	2.96	1.02
Pendleton	74-76	64.46	0.99	1.01	62.55	0.95	0.96	1.58	0.45
Red Bluff	54-56	31.70	-0.64	0.72	25.41	-0.55	0.64	1.93	0.62
Red Bluff	74-76	43.94	-0.99	1.08	39.03	-0.89	0.99	2.19	0.70
Salem	54-56	26.10	-0.79	0.95	24.27	-0.75	0.95	2.73	0.99
Salem	74-76	4.32	0.04	0.22	3.67	-0.09	0.16	3.14	0.64
Yuma	74-76	11.06	0.14	0.14	18.75	0.22	0.23	1.22	0.28
For the years 1954-1956		24.26	-0.81	1.04	24.88	-0.91	1.12	3.54	1.63
For the years 1974-1976		26.58	-0.38	1.16	28.09	-0.50	1.41	3.20	2.54
For all the years		25.65	-0.55	1.11	26.80	-0.66	1.30	3.34	2.23
Rocky Mountain									
Billings	74-76	42.43	0.34	0.35	54.03	0.44	0.45	0.92	0.35
Colorado Springs	74-76	24.31	-0.26	0.27	14.10	-0.15	0.15	1.10	0.22
Del Rio	73-75	7.22	-0.10	0.16	5.06	-0.08	0.11	2.19	0.38
Farmington	74-76	22.10	-0.19	0.23	8.17	-0.09	0.15	0.82	0.27
Grand Junction	54-56	34.20	0.22	0.23	51.19	0.33	0.34	0.68	0.19
Grand Junction	74-76	15.33	-0.01	0.19	24.67	0.10	0.22	0.93	0.41
Kalispell	74-76	3.56	0.04	0.06	7.48	0.10	0.12	1.70	0.44
Rock Springs	74-76	4.93	-0.03	0.07	10.04	0.07	0.10	1.04	0.30
Tucson	54-56	73.85	0.50	0.52	89.47	0.60	0.62	0.66	0.12
Tucson Intl	74-76	118.31	0.73	0.76	133.92	0.82	0.84	0.60	0.09
For the years 1954-1956		54.02	0.36	0.40	70.33	0.47	0.50	0.67	0.16
For the years 1974-1976		29.77	0.06	0.33	32.18	0.15	0.36	1.16	0.59
For all the years		34.62	0.12	0.35	39.81	0.21	0.39	1.06	0.56
Central									
Dubuque	74-76	8.99	-0.21	0.35	11.44	-0.40	0.42	3.71	0.55
Evansville	54-56	41.39	1.72	1.83	26.22	1.09	1.15	4.04	0.72
Evansville	74-76	49.04	2.28	2.40	30.32	1.40	1.47	4.53	0.83
Fort Smith	74-76	8.45	0.29	0.41	5.44	0.13	0.23	3.03	0.49
Huron	74-76	23.48	0.68	0.76	17.49	0.50	0.55	2.77	0.47
Jackson	74-76	62.65	2.20	2.39	45.56	1.60	1.74	3.40	0.38
Port Arthur	74-76	51.83	1.44	1.57	41.17	1.14	1.23	2.68	0.35
Sault St Marie	74-76	42.80	1.32	1.40	32.08	0.98	1.02	3.01	0.56
For the years 1954-1956		41.39	1.72	1.83	26.22	1.09	1.15	4.04	0.72
For the years 1974-1976		35.32	1.14	1.55	26.21	0.77	1.09	3.30	0.81
For all the years		36.08	1.21	1.59	26.21	0.81	1.09	3.40	0.83
Eastern									
Augusta	54-56	60.20	1.75	1.90	47.07	1.37	1.47	2.81	0.34
Augusta	74-76	57.31	2.48	2.67	37.82	1.64	1.77	4.18	0.55
Binghampton	54-56	5.95	-0.15	0.21	9.57	-0.33	0.33	3.57	0.65
Binghampton	74-76	18.29	-0.83	0.89	22.52	-1.03	1.12	4.41	1.13
Burlington	74-76	14.99	-0.52	0.58	17.12	-0.60	0.70	3.39	0.87
Dulles	74-76	4.26	-0.13	0.20	10.80	-0.47	0.48	4.44	0.88
Roanoke	54-56	8.47	-0.21	0.28	10.60	-0.33	0.34	3.29	0.54
Roanoke	74-76	40.94	-1.46	1.47	40.40	-1.44	1.45	3.61	0.62
Worcester	74-76	23.91	-1.04	1.05	37.42	-1.21	1.23	4.44	0.59
Youngstown	54-56	6.14	0.31	0.33	4.79	-0.28	0.35	5.26	1.25
Youngstown	74-76	12.46	0.74	0.83	1.01	-0.04	0.08	5.65	1.14
For the years 1954-1956		20.19	0.43	0.98	18.01	0.11	0.79	3.73	1.21
For the years 1974-1976		24.39	-0.11	1.32	22.44	-0.45	1.12	4.30	1.12
For all the years		22.99	0.09	1.21	20.83	-0.25	1.01	4.10	1.18
Nation-wide									
For the years 1954-1956		29.75	0.08	1.04	30.77	-0.11	0.92	3.73	1.21
For the years 1974-1976		29.18	0.20	1.17	27.38	0.02	1.03	4.30	1.12
For all the years		29.34	0.16	1.13	28.33	-0.02	1.00	4.10	1.18

*SD = standard deviation.

regression relationship tended to fit the data about as well in the Rocky Mountain Region.

By examining the AAPE statistic, one can note differences in prediction performance between regions and quarters. Except for the national ratio relationship, the relationships perform best in the Central and Eastern Regions and worst in the Rocky Mountain Region. The relationships also perform best on an annual basis, and for the second quarter, and perform worst for the first, third, and fourth quarters.

Locations with large errors according to the AAPE statistic (greater than 50%) were identified. Pendleton (1974-1976) and Tucson (1954-1956, 1974-1976) had large errors given by all four relationships on an annual and a quarterly basis. The regression prediction errors were large for Grand Junction (1954-1956, 1974-1976), particularly for the second through fourth quarters. All relationships, except the national regression relationship, performed poorly for Augusta (1954-1956, 1974-1976).

The recommended relationships for making future forecasts of upper percentile visibility from a forecast of median visibility for locations in the United States were selected according to their simplicity and their performance on the test data set. The regional ratio relationship is recommended for annual, first, and fourth quarter forecasts. The national ratio relationship is recommended for second and third quarter forecasts.

5. SUMMARY AND CONCLUSIONS

Statistical methods have been used to estimate and assess the performance of relationships between median and upper percentile (worst-case) extinction. These relationships extend a methodology developed by LASL for estimating median regional scale visibility. With the relationship between median and upper percentile extinction, regional scale visibility for worst-case days can be estimated. Airport weather station measurements of visual range were used in the estimation and assessment of the relationships. Airport weather stations located in the continental US were screened for their reporting practices, distances to farthest marker, site relocations, and their representativeness of regional visibility. Twenty-eight suburban/nonurban sites that met the screening criteria were selected. The data were sorted according to meteorology in order to eliminate days that are obviously dominated by natural causes (e.g. fog, precipitation, blowing dust, or low clouds) of poor visibility. The Koschmeider relationship was used to translate the visual range data to extinction data. Three approaches were followed to develop relationships between median and upper percentile light extinction. One is based upon frequency distribution functions. The second uses observed ratios of upper percentile to median extinction. The third employs regression techniques. The average absolute percent errors for the three approaches predominantly ranged between 20-40%, which compares well with air quality models that frequently give factor-of-two agreement with observation. Simple ratio relationships are recommended for use in translating median visibility impacts into worst-case impacts.

Acknowledgement - This paper is based on work performed under contract with the US Department of Energy, Office of the Assistant Secretary for the Environment, Office of Technology Impacts, Regional Assessments Division.

REFERENCES

Breiman, L., (1973) Statistics: With a View Toward Applications, Houghton Mifflin, Boston.

Husar, R.B., Patterson, D.E., and Holloway, J.M., (1979) Trends of Eastern U.S. haziness since 1948, Fourth Symp. on Atmospheric Turbulence Diffusion, and Air Pollution, Reno, Nevada, January 15-18, 1979.

Latimer, D., Bergstrom, R.W., Hayes, S.R., Liu, M.K., Seinfeld, J.H., Whitten, G.Z., Wojcik, M.A., and Hillyer, M.J. (1978) The development of mathematical models for the prediction of anthropogenic visibility impairment, EPA-450/3/78 110a, US Environmental Protection Agency, Research Triangle Park, North Carolina.

Marians, M. and Trijonis, J., (1979) Empirical studies of the relationship between emissions and visibility in the Southwest, Prepared at Technology Service Corporation for EPA Office of Air Quality Planning and Standards, Atmospheric Chemistry and Physics Division, Research Triangle Park, North Carolina.

Nochumson, D.H., Wecksung, W., and Gurule, F., (1979) National impacts on visual air quality from a future energy scenario, Proc. View on Visibility-Regulatory and Scientific, sponsored by the Air Pollution Control Association, Denver, Colorado, November 26-28, 1979.

Nochumson, D. H., Gurule, F., and Wecksung, M.J. (1979) Potential future impacts on visual air quality for class I areas, Proc. of Our National Landscape--A Conference on Applied Techniques for Analysis and Management of the Visual Resource, Incline Village, Nevada, April 23-25, 1979.

Roberts, F.M., Gordon, J.L., and Haase, D.L., (1975) Visibility Measurements in the Painted Desert, Proc. at the 68th Annual Meeting of the Air Pollution Control Association, Boston, Massachusetts, June 15-20, 1975.

Trijonis, T. and Yuan, K., (1978) Visibility in the Northeast: Long-term visibility trends and visibility/pollutant relationships, EPA-600/3-78-075 US Environmental Protection Agency, Research Triangle Park, North Carolina.

Trijonis, T., (1979) Visibility in the Southwest--An Exploration of the Historical Data Base, in Atmospheric Environment, Vol. 13, p. 833.

Trijonis, T. and Shapland, D., (1979) Existing visibility levels in the US--Isopleth maps of visibility in suburban/nonurban areas during 1974-1976, Prepared at Technology Service Corporation for EPA Office of Air Quality Planning and Standards and EPA Atmospheric Chemistry and Physics Division, Research Triangle Park, North Carolina.